

# SUPPORTING INFORMATION

Nucleic Acids Research, 2014

doi: [10.1093/nar/gku855](https://doi.org/10.1093/nar/gku855)

## **μABC: a systematic microsecond molecular dynamics study of tetranucleotide sequence effects in B-DNA**

Marco Pasi<sup>1</sup>, John H. Maddocks<sup>1,\*</sup>, David Beveridge<sup>2</sup>, Thomas C. Bishop<sup>3</sup>, David A. Case<sup>4</sup>, Thomas Cheatham III<sup>5</sup>, Pablo D. Dans<sup>6</sup>, B. Jayaram<sup>7</sup>, Filip Lankas<sup>8</sup>, Charles Laughton<sup>9</sup>, Jonathan Mitchell<sup>1</sup>, Roman Osman<sup>10</sup>, Modesto Orozco<sup>6</sup>, Alberto Pérez<sup>6</sup>, Daiva Petkevičiūtė<sup>1</sup>, Nada Spackova<sup>11</sup>, Jiri Spomer<sup>11,12</sup>, Krystyna Zakrzewska<sup>13</sup> and Richard Lavery<sup>13,\*</sup>

<sup>1</sup> Section de Mathématiques, Swiss Federal Institute of Technology (EPFL), CH-1015 Lausanne, Switzerland

<sup>2</sup> Department of Chemistry, Wesleyan University, Middletown, CT 06459, USA

<sup>3</sup> Departments of Chemistry & Physics, Louisiana Tech University, Ruston, LA 71270, USA

<sup>4</sup> BioMaPS Institute and Dept. of Chemistry & Chemical Biology, Rutgers University, 610 Taylor Road, Piscataway, NJ 08854-8087, USA

<sup>5</sup> Department of Medicinal Chemistry, University of Utah, Skaggs 307, Salt Lake City UT. 84112, USA

<sup>6</sup> Joint BSC-CRG-IRB Program on Computational Biology, Institute of Research in Biomedicine, Parc Científic de Barcelona, Josep Samitier 1-5, Barcelona 08028. Spain and Barcelona Supercomputing Centre, Jordi Girona 31, Edifici Torre Girona. Barcelona 08034, and Departament de Bioquímica, Facultat de Biología, Avgda Diagonal 647, Barcelona 08028, Spain

<sup>7</sup> Department of Chemistry, Indian Institute of Technology, Hauz Khas, New Delhi 110016, India

<sup>8</sup> Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo nam. 2, 166 10 Praha 6, Czech Republic

<sup>9</sup> School of Pharmacy and Centre for Biomolecular Sciences, University of Nottingham, NG7 2RD, UK

<sup>10</sup> Department of Structural and Chemical Biology, Mount Sinai School of Medicine, New York, NY 10029, USA

<sup>11</sup> Institute of Biophysics, Academy of Sciences of the Czech Republic, Kralovopolska 135, 612 65 Brno, Czech Republic

<sup>12</sup> CEITEC - Central European Institute of Technology, Masaryk University, Campus Bohunice, Kamenice 5, 625 00 Brno, Czech Republic

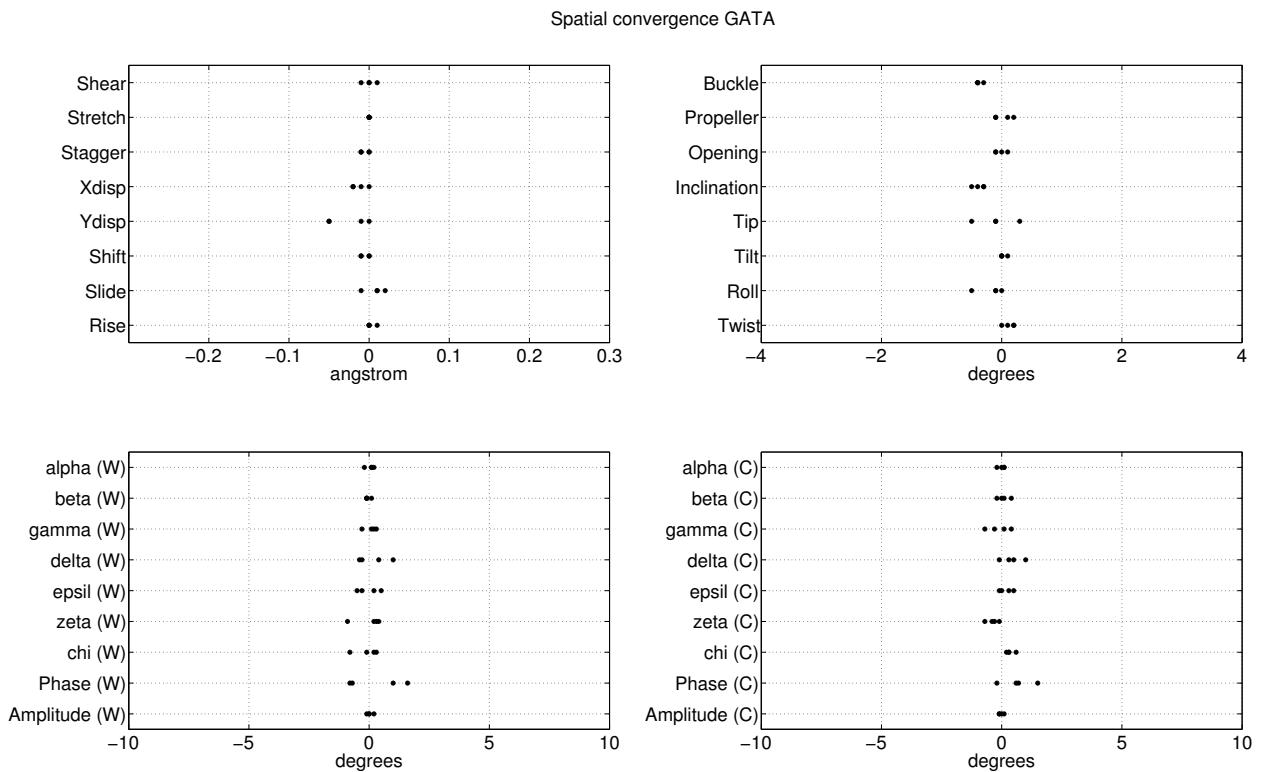
<sup>13</sup> Bases Moléculaires et Structurales des Systèmes Infectieux, CNRS UMR 5086 / Univ. Lyon I, IBCP, 7 Passage du Vercors, 69367 Lyon, France

\*To whom correspondence should be addressed. Email: [richard.lavery@ibcp.fr](mailto:richard.lavery@ibcp.fr); Tel: +33 4 72 72 26 37; Fax +33 4 72 72 26 04. Email: [john.maddocks@epfl.ch](mailto:john.maddocks@epfl.ch); Tel: +41 21 693 27 62; Fax: +41 21 693 55 30.

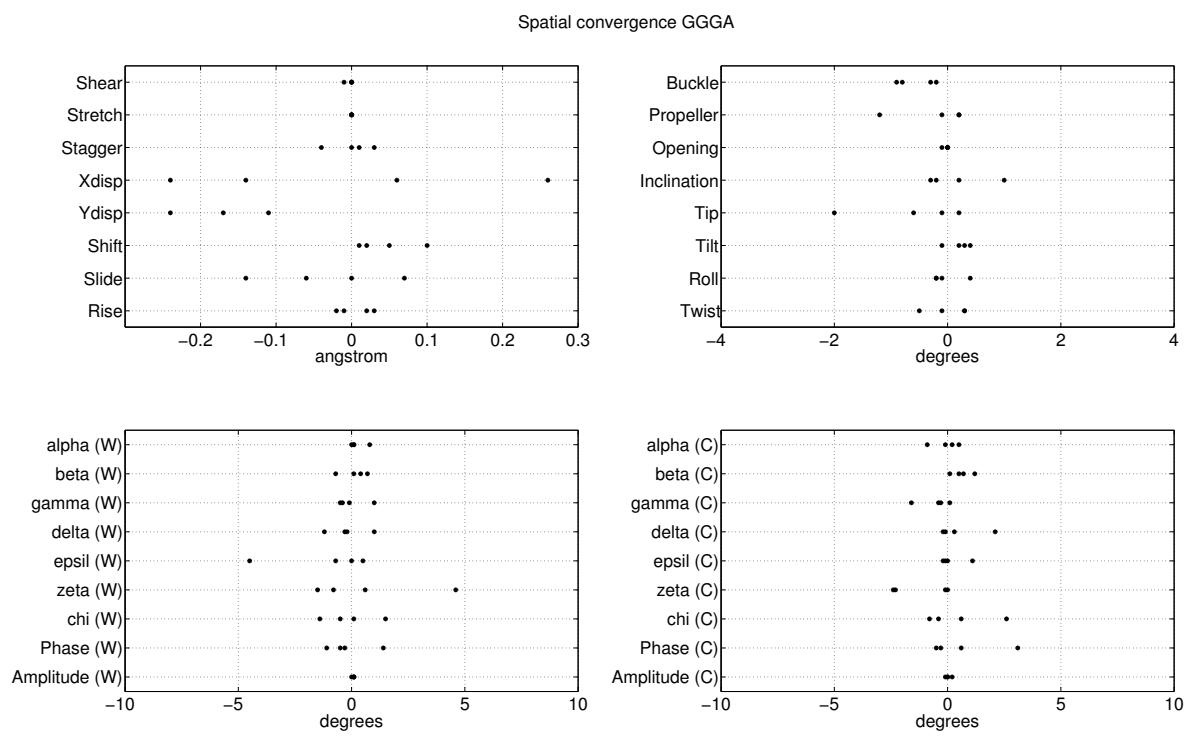
Table S1. The 39 18-base pair oligomers studied in this work, showing the unique tetranucleotides constituting the repeating sequence of each oligomer.

Name	Sequence	Tetranucleotides
AAAA	GCAAAAAAAAAGC	AAAA
AAAC	GCACAAACAAACAAACGC	AAAC, AACA, ACAA, CAAA
AAAT	GCATAAATAAAATAATGC	AAAT, AATA, ATAA, TAAA
AGAG	GCAGAGAGAGAGAGAGGC	AGAG, GAGA
AGCG	GCCGAGCGAGCGAGCGGC	AGCG, GCGA, CGAG, GAGC
AGCT	GCCTAGCTAGCTAGCTGC	AGCT, GCTA, CTAG
AGGA	GCGAAGGAAGGAAGGAGC	AGGA, GGAA, GAAG, AAGG
AGGC	GCGCAGGCAGGCAGGCGC	AGGC, GGCA, GCAG, CAGG
AGGT	GCGTAGGTAGGTAGGTGC	AGGT, GGTA, GTAG, TAGG
AGTC	GCTCAGTCAGTCAGTCGC	AGTC, GTCA, TCAG, CAGT
AGTG	GCTGAGTGAGTGAGTGGC	AGTG, GTGA, TGAG, GAGT
ATGC	GCGCATGCATGCATGCGC	ATGC, TGCA, CATG
CAAG	GCAGCAAGCAAGCAAGGC	CAAG, AAGC, AGCA, GCAA
CAAT	GCATCAATCAATCAATGC	CAAT, AATC, ATCA, TCAA
CGCG	GCCGCGCGCGCGCGCGC	CGCG, GCGC
CGGA	GCGACGGACGGACGGAGC	CGGA, GGAC, GACG, ACGG
CGGC	GCGCCGGCCGGCCGGCGC	CGGC, GGCC, CCGG
CGGT	GCGTCGGTCGGTCGGTGC	CGGT, GGTC, GTCG, TCGG
CGTA	GCTACGTACGTACGTAGC	CGTA, GTAC, ACGT
CGTG	GCTGCGTGCCTGCCTGGC	CGTG, GTGC, TGCG, GCGT
GAAA	GCAAGAAAGAAAGAAAGC	GAAA, AAAG, AAGA, AGAA
GAAC	GCACGAACGAACGAACGC	GAAC, AACG, ACGA, CGAA
GAAT	GCATGAATGAATGAATGC	GAAT, AATG, ATGA, TGAA
GATA	GCTAGATAGATAGATAGC	GATA, ATAG, TAGA, AGAT
GGGA	GCGAGGGAGGGAGGGAGC	GGGA, GGAG, GAGG, AGGG

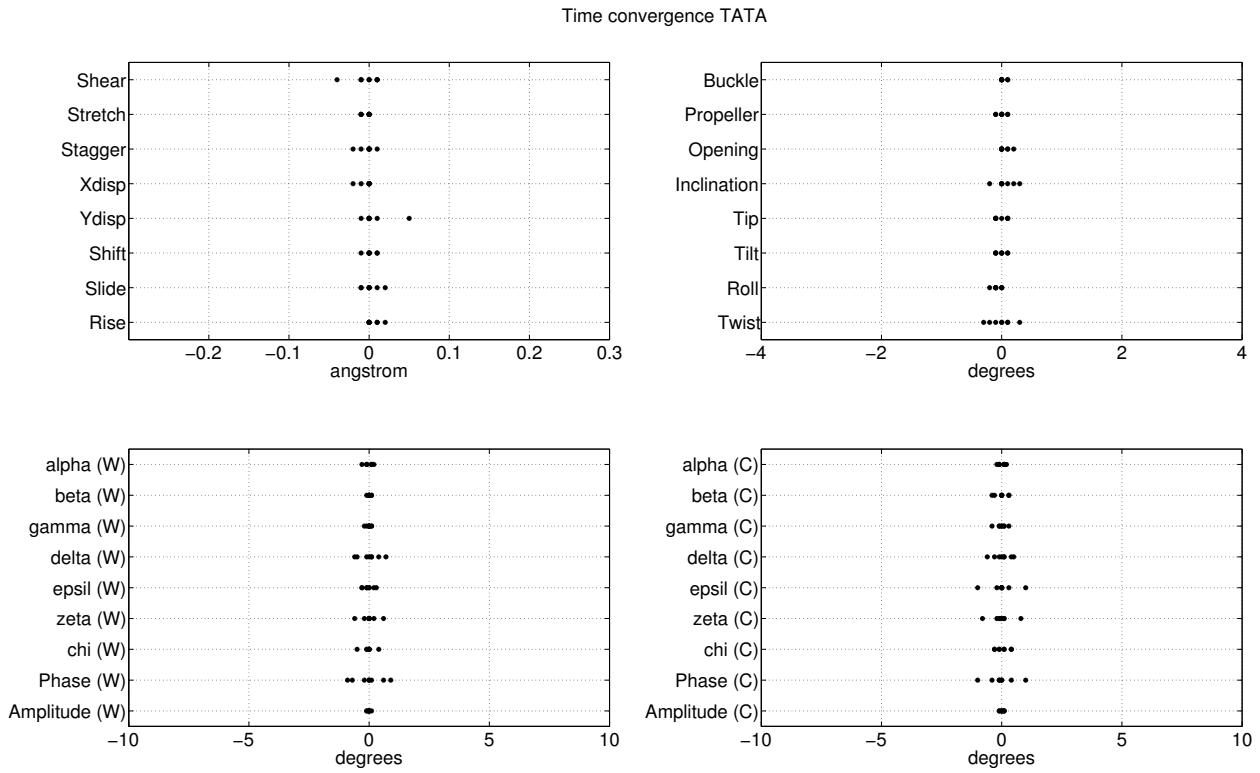
Name	Sequence	Tetranucleotides
GGGC	GCGCGGGCGGGCGGGCGC	GGGC, GGCG, GCGG, CGGG
GGGG	GCGGGGGGGGGGGGGGGC	GGGG
GGGT	GCGTGGGTGGGTGGTGC	GGGT, GGTG, GTGG, TGGG
TAAC	GCACTAACTAACTAACGC	TAAC, AACT, ACTA, CTAA
TAAG	GCAGTAAGTAAGTAAGGC	TAAG, AAGT, AGTA, GTAA
TAAT	GCATTAATTAAATTAATGC	TAAT, AATT, TTAA
TATA	GCTATATATATATATAGC	TATA, ATAT
TCGA	GCGATCGATCGATCGAGC	TCGA, CGAT, GATC
TGGA	GCGATGGATGGATGGAGC	TGGA, GGAT, GATG, ATGG
TGGC	GCGCTGGCTGGCTGGCGC	TGGC, GGCT, GCTG, CTGG
TGGT	GCGTTGGTTGGTTGGTGC	TGGT, GGTT, GTTG, TTGG
TGTA	GCTATGTATGTATGTAGC	TGTA, GTAT, TATG, ATGT
TGTC	GCTCTGTCTGTCTGTCGC	TGTC, GTCT, TCTG, CTGT
TGTG	GCTGTGTGTGTGTGGC	TGTG, GTGT



**Figure S1: Conformational convergence of  $\mu$ ABC simulations.** Spatial convergence of the simulations for a chosen ABC oligomer was assessed by comparing average helical and backbone parameters within two occurrences of the same tetranucleotide sequence. The two tetranucleotides were chosen at positions 6-9 and 10-13, towards the center of the oligomer. The differences of the averages are plotted as black dots for each base pair or base pair step within the tetranucleotides. This figure shows the results for the GATA oligomer, which was amongst those with the best conformational convergence properties.

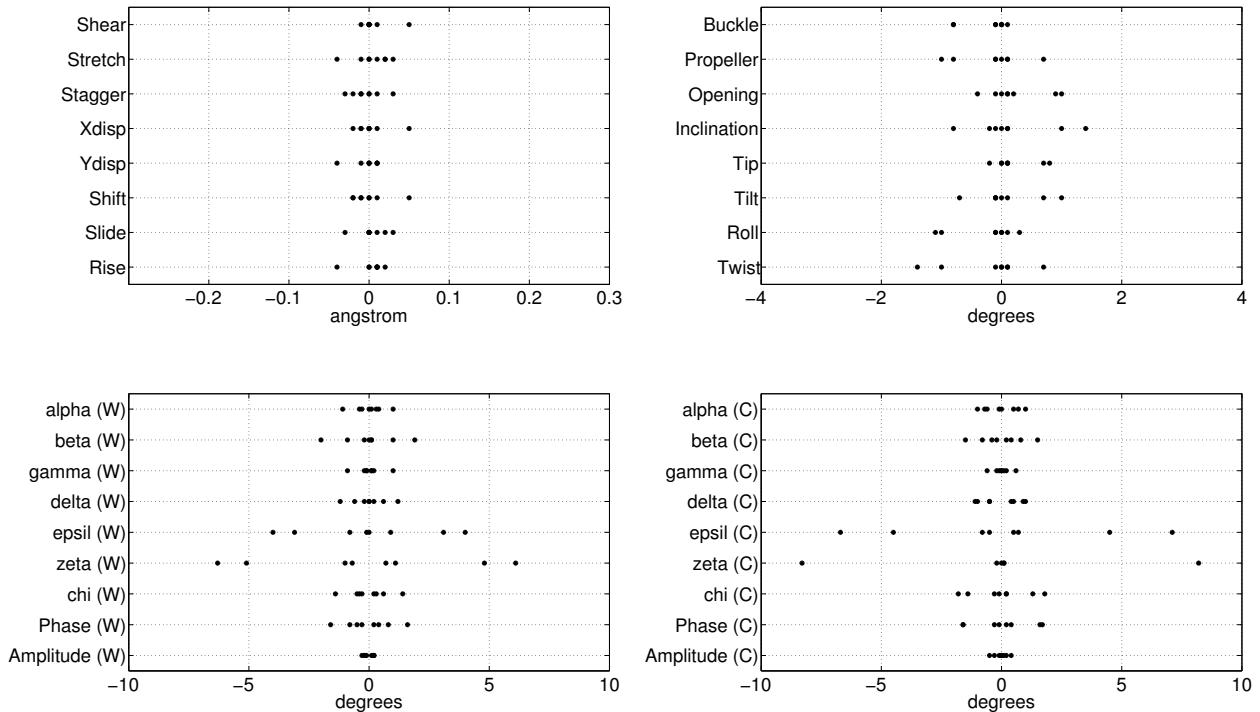


**Figure S2: Conformational convergence of  $\mu$ ABC simulations.** Results for the GGGA oligomer, which was amongst those with the worst conformational convergence properties (see figure S1 for details)

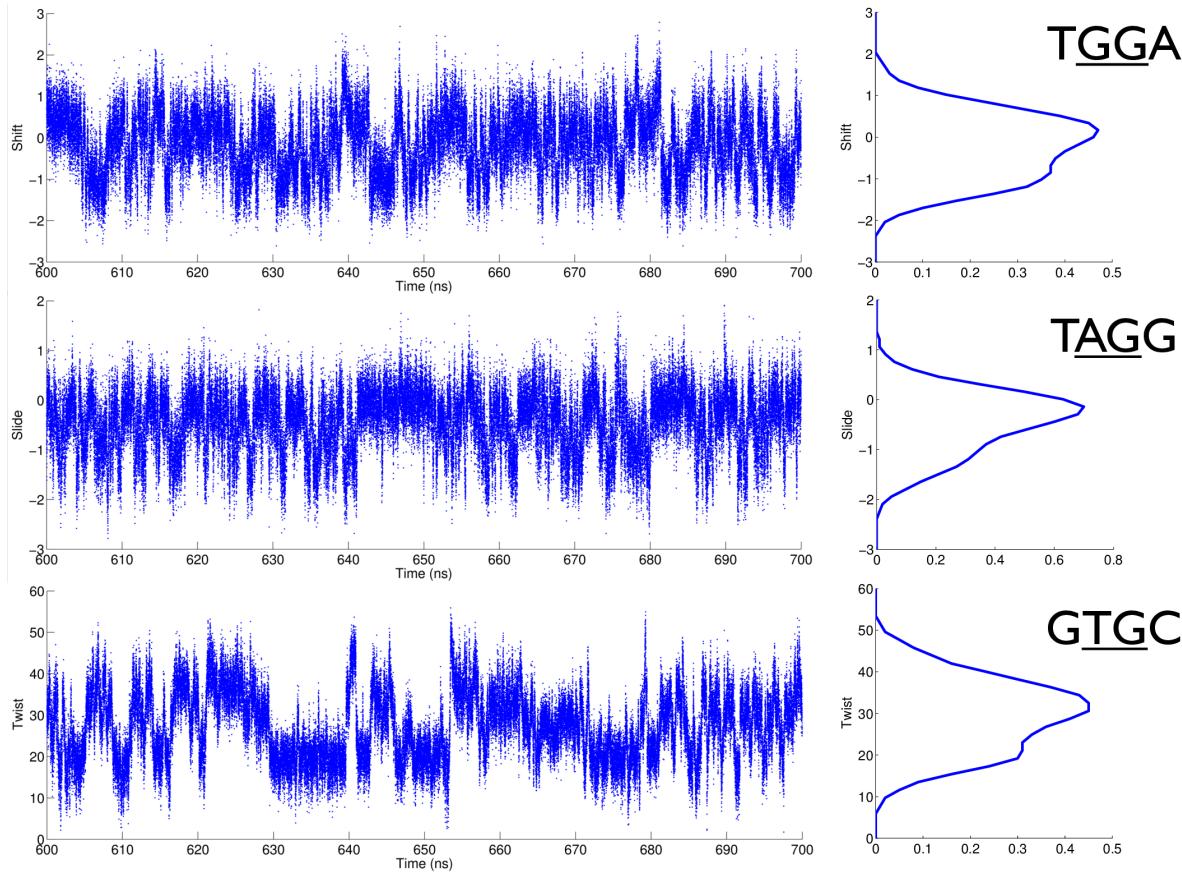


**Figure S3: Time convergence of  $\mu$ ABC simulations.** Time convergence of simulations for a chosen ABC oligomer was assessed by comparing average helical and backbone parameters calculated using either the first half ( $H_1$ ) or the second half ( $H_2$ ) of the corresponding trajectory with the averages obtained for the entire trajectory (T). The differences  $T-H_1$  and  $T-H_2$  are plotted as black dots for each base pair or base pair step within the central tetranucleotide (positions 7-10). This figure shows the results for the TATA oligomer, which was amongst those with the best time convergence properties.

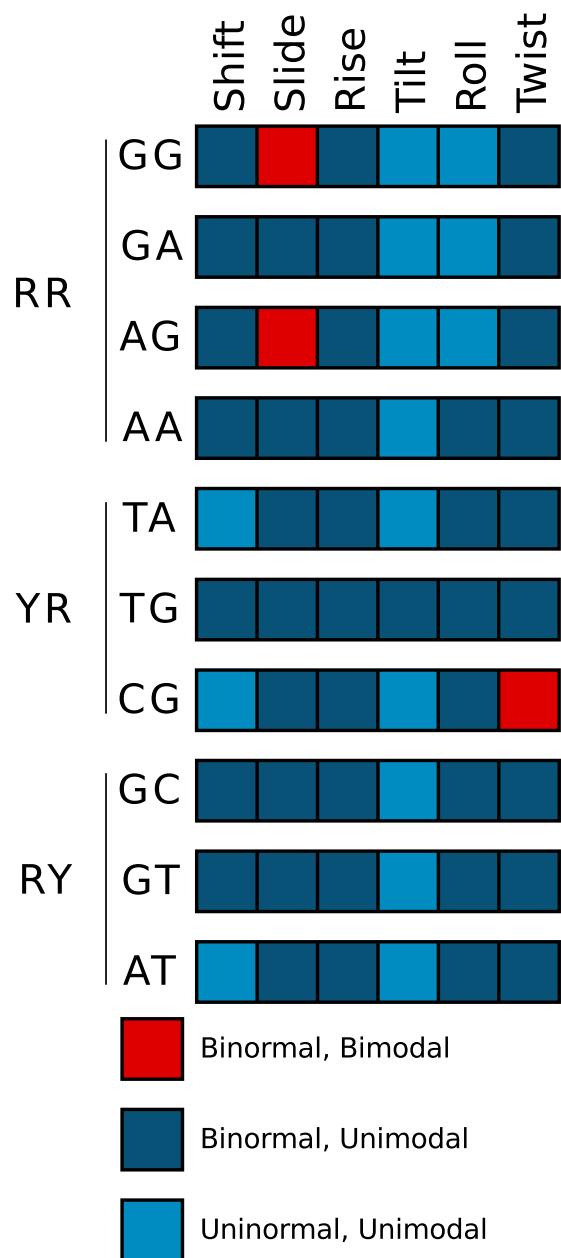
Time convergence CGCG



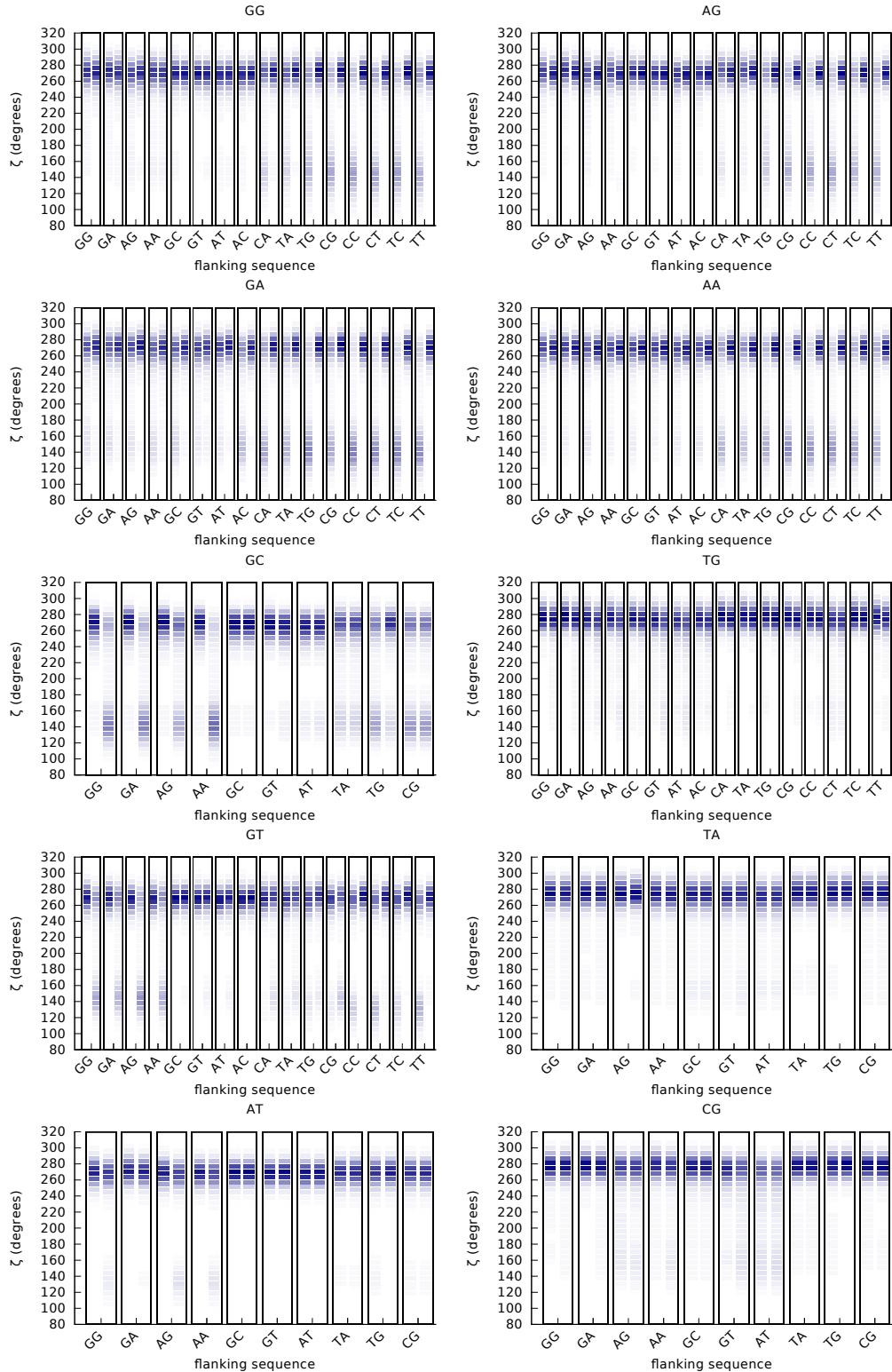
**Figure S4: Time convergence of  $\mu$ ABC simulations.** Results for the CGCG oligomer, which was amongst those with the worst time convergence properties (see Figure S3 for details).



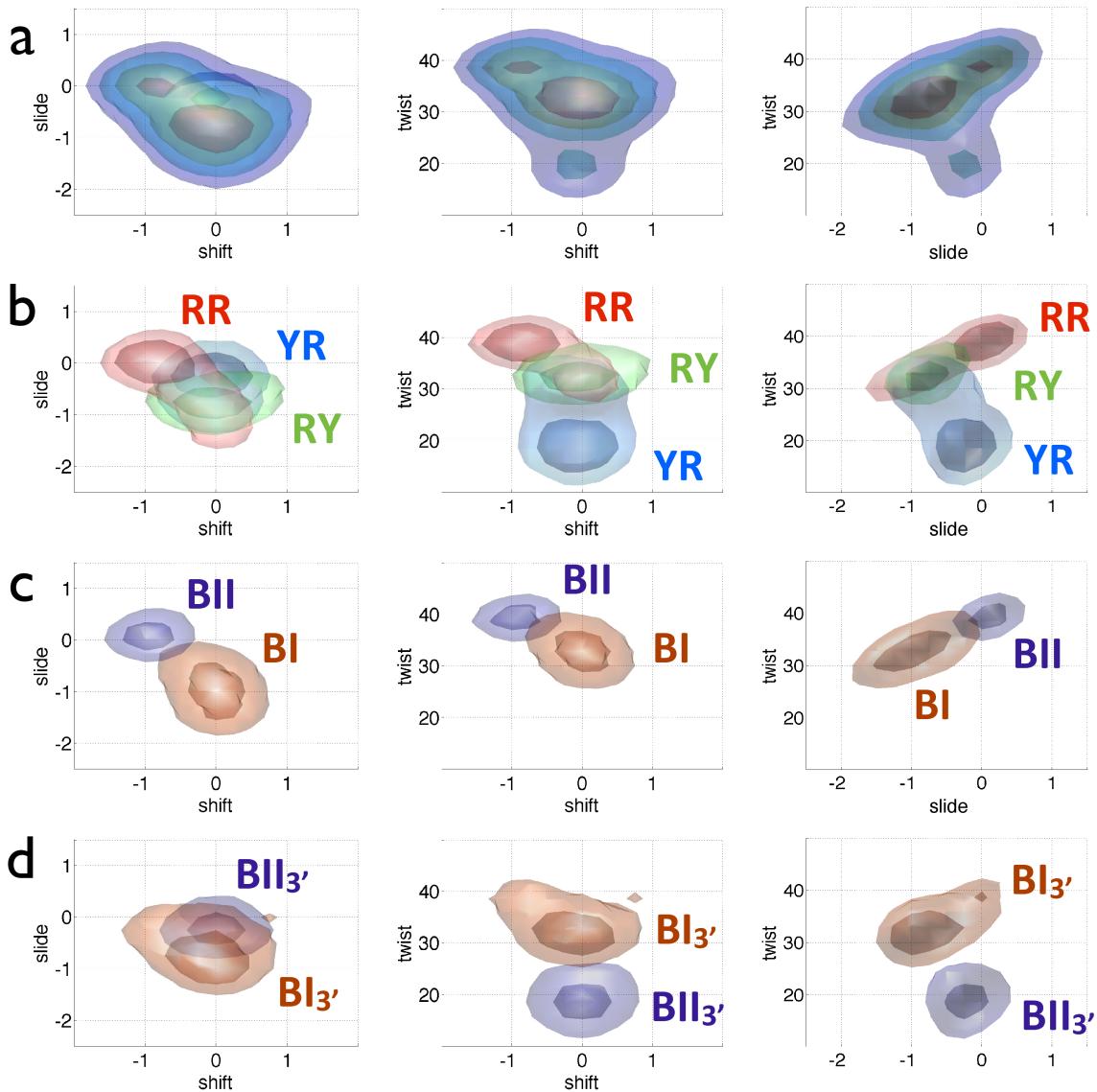
**Figure S5: Evidence of conformational transitions.** Three examples of inter-BP parameter time series (with the corresponding time-averaged distributions) showing transitions between at least two conformational substates: TGGA shift (top), TAGG slide (middle); GTGC twist (bottom). The parameters refer to the tetranucleotide closest to the center of each oligonucleotide and, for clarity, are limited to a 100 ns interval extracted after 600 ns of simulation.



**Figure S6: Deviations from Gaussianity in helical parameter distributions.** This presentation has been adapted from results in by Dans *et al.* (*Nucleic Acids Res.* 2012 40:10668-10678) for ease of comparison with the data shown in Figure 4 of the main text.



**Figure S7: Tetranucleotide sequence effects on backbone dihedral angle  $\zeta$ .** The distribution of the backbone dihedral angle  $\zeta$  is computed for both strands of the central base pair step of each of the 136 unique tetranucleotide sequences. Distributions are shown as histograms for the Watson and Crick strands (left and right column respectively inside the black frames), where each small rectangle represents a  $6.5^\circ$  bin, and the color intensity is proportional to the BII population. For each central base pair step (one per panel), either 16 pairs or (for self-complementary steps) 10 pairs of histograms are plotted grouped according to flanking sequence X..Y (shown along the abscissa).



**Figure S8: Two-dimensional projections of shift-slide-twist probability distributions.** Two-dimensional projections shift-slide (left column), shift-twist (middle column), and slide-twist (right column) of the four probability density plots shown in figure 8: the rows from top to bottom are the three 2D projections of, respectively, panels a to d.